DOCUMENT RESUME

ED 392 841 TM 024 667

AUTHOR	Gadalla, Tahany
TITLE	A Confirmatory Factor Analysis and a Group Comparison Analysis of the IEA Student Attitude Questionnaire.
PUB DATE	Apr 95
NOTE	14p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Comparative Analysis; Cross Cultural Studies; *Factor Structure; Foreign Countries; *Goodness of Fit; International Education; Models; Questionnaires; Sex Differences; *Student Attitudes; *Test Items; Test Reliability; Test Use; Test Validity
IDENTIFIERS	Canada; Confirmatory Factor Analysis; *International Assn Evaluation Educ Achievement; Missing Data; *Second International Mathematics Study

ABSTRACT

The Second International Mathematics Study was conducted in 20 countries under the sponsorship of the International Association for Evaluation of Educational Achievement (IEA). Among the instruments used in this study was a questionnaire investigating student attitudes about school, instruction, and mathematics. The fit of an a priori model that postulates the relationships between observed responses to sets of items comprising different scales and the latent traits the scales are designed to indicate was studied. The hypothesis of equality of factor structure for boys and girls was also tested. Data were from 13-year-old Canadian students, 2,422 boys and 2,401 girls. The PRELIS computer program was used to compute the polychoric correlations and the asymptotic covariance matrices for the response data, using listwise deletion of missing responses. Analyses made it apparent that most items needed revision, that they had poor reliability, and that they seemed to be measuring a mixture of traits. Insufficient evidence was found concerning the hypothesis that boys and girls had the same factor pattern. The validity of international comparisons from survey responses has not been established. (Contains two figures and three tables.) (SLD)

A Confirmatory Factor Analysis and A Group Comparison Analysis of the IEA Student Attitude Questionnaire

Tahany Gadalla

Ontario Institute For Studies in Education

Presented at the 1995 AERA Annual Meeting

San Francisco, California, April 18-22

I. Introduction

The Second International Mathematics Study (SIMS) was conducted in the schools of 20 education systems under the sponsorship of the International Association for Evaluation of Educational Achievement (IEA). Among the various instruments used in this study is a questionnaire which includes over 90 items which were designed to investigate students' thoughts about various aspects of schools, instructions, and mathematics. The items solicited responses regarding attitudes, beliefs, and opinions related to the study of mathematics. The items are divided into 7 sets, each set comprising a scale which is used to measure a specific trait. The scales are: home support, mathematics in schools, mathematics as a process, mathematics and myself, mathematics anxiety, gender stereotyping, and utility of mathematics. Responses to all items were measured on a 5-point Likert scale of the format:

Strongly agree   agree   Undecided   Disagree   Strongly disagree

II. Objectives

The purpose of this paper is twofold. First, to examine the fit of the a priori model that postulates the relationships between observed responses to sets of items comprising different scales, and the latent traits which the scales are designed to indicate. In other words, to describe how well the items serve as measuring instruments for their hypothesized latent traits. If the model does not fit the data well, an attempt will be made at finding an alternative model that fits the data better. Second, to test the hypothesis of equality of factor structures for boys and girls. This entails the testing of the hypotheses of similar factor patterns, equal units of measurement, and equal accuracy of measurement for the two groups.

III. Data

The data used in the present study were collected from 13 year old Ontario students during the period 1980-82. The data file contained sets of responses from 4823 students, of whom 2422 were boys and 2401 were girls.

Due to the limitations on the memory size in PRELIS, only 3 of the scales used in the questionnaire are investigated. A total of 17 items comprise the subject of this investigation. The following scales are selected for study.

Mathematics Utility: This scale addressed students perceptions of the practicality and usefulness of mathematics in everyday life. The scale is comprised of the following 8 items:

mu1. It is important to know math to get a good job.

mu2. Most people do not use math in their jobs.

mu3. I would like a job that lets me use math.

mu4. Math is useful in solving everyday problems.

mu5. I can get along well in everyday life without math.

mu6. Most math has practical use on the job.

mu7. Math is not needed in everyday living.

mu8. A knowledge of math is not necessary in most occupations.

Mathematics anxiety: This scale was intended to measure the extent to which students find dealing with mathematics unsettling or frightening. This scale consists of 5 items:

ma1. Working with numbers makes me happy.

ma2. It scares me to have to take math.

ma3. I usually feel calm when doing math problems.

ma4. I think math is fun.

ma5. When I cannot figure out a math problem I feel lost in a maze.

Gender stereotyping: Four items were designed to tap into a student's gender stereotyping attitude towards mathematics:

ms1. Men make better scientists and engineers than women.

ms2. Boys have more natural ability in math than girls.

ms3. Boys need to know more math than girls.

ms4. A woman needs a career just as much as a man does.

IV. Methods

PRELIS program is used to compute the polychoric correlations and the asymptotic covariance matrices for the response data, using listwise deletion of missing responses. Three analyses are carried out.

Analysis 1. Confirmatory factor analysis is used to test the goodness of fit of the model proposed by the IEA officials, and to estimate the reliability of the items in measuring the traits which they are intended to measure. This is carried out on boys and girls data separately.

Analysis 2. The sample of boys is divided in half. The odd-numbered cases are used to develop the model, i.e. the training data set, and the even-numbered cases are used to test the model. Girls responses are divided in half in the same manner. A LISREL measurement model describing the observed relationships between the items and the latent traits is developed for the boys training data set by means of weighted least squares estimation procedure. This model is tested on the other half of the boys data in an attempt to make sure that the model does not capitalize on the peculiarities of the training set.

Analysis 3. Once the boys model is finalized, a multisample LISREL analysis is carried out in order to test the invariance of item functioning for boys and girls. This is done in 3 steps.

Step 1. To test the hypothesis that both boys and girls have the same factor patterns.

Step 2. Given that the two groups have the same factor pattern, to test the hypothesis that the corresponding factor loadings are equal.

Step 3. Given that the above two hypotheses are true, to test the hypothesis that the corresponding latent traits are measured with the same accuracy for both groups. In other words, to test the hypothesis that the standard errors of the factor loadings for the two groups are equal.

V. Results

Analysis 1.

A measurement model of the form $X = \Lambda\xi + \delta$,

where $X' = (X_1, \ldots, X_{17})$ are the observed variables,

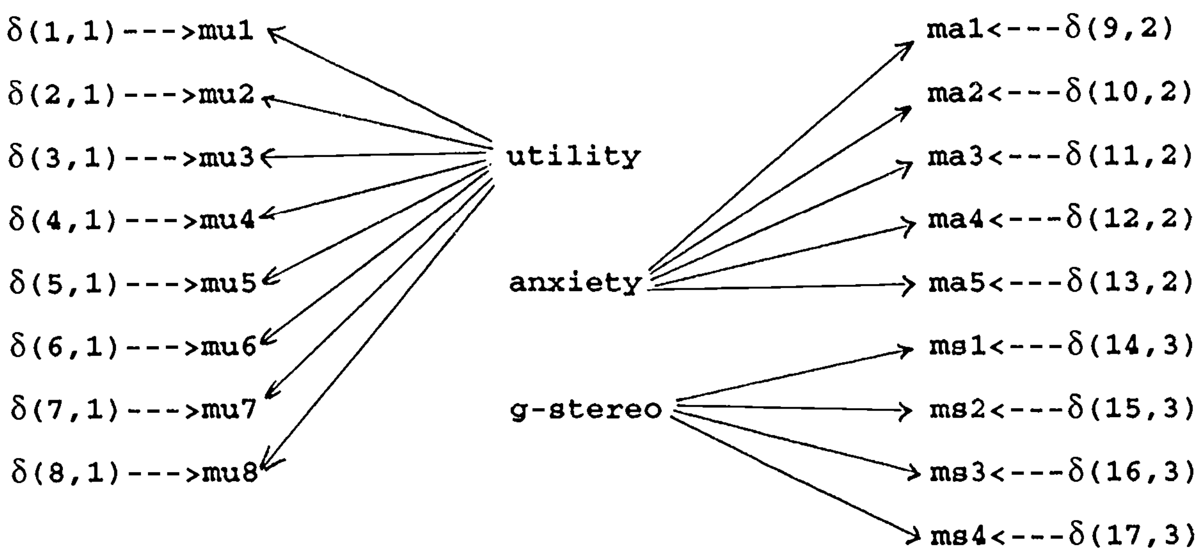
$\xi' = (\xi_1, \xi_2, \xi_3)$ are the latent variables, and

$\delta' = (\delta_1, \ldots\ldots, \delta_{17})$ are the error terms,

was fit to each data set.

The expected variance-covariance matrix of the items is $\Sigma = \Lambda\Phi\Lambda' + \Theta$, where $\Phi(3x3)$ is the variance-covariance matrix of the latent variables, and $\Theta(17x17)$ is a diagonal variance-covariance matrix of the errors.

Figure 1. Postulated model.



The measurement model postulated by the IEA officials is shown in Figure 1, where mu1 to mu8 refer to the eight items on the math utility scale, ma1 to ma5 refer to the five items on the math anxiety scale, and ms1 to ms4 refer to the four items on the math stereotyping scale. The unit of measurement for each latent factor is determined by the first item on its scale. This is done by fixing $\Lambda(1,1)$, $\Lambda(9,2)$, and $\Lambda(14,3)$ to unity. The model is fit to the boys' response data as well as the girls'.

Table 1. Goodness of fit of the a priori model - Analysis 1.

	Boys data	Girls data
χ^2 index	7.58	9.59
Range of standardised residuals	-6.12 , 11.61	-9.06 , 10.10

Table 1 shows some measures of goodness of fit of the postulated model. The high χ^2 indices for both boys and girls and the large standardised residuals indicate rather poor fits.

Table 2 shows the multiple squared correlations calculated by LISREL, which measure the reliability of each item. As is apparent from table 2, most of the items have poor reliability in measuring the latent traits which they are intended to measure. Reliability estimates for the 8 measures of "math utility" varied between 0.20 and 0.39, with the exception of item no. 7 for girls, for which the reliability was estimated at 0.46. Reliability estimates for the first 4 measures of "math anxiety" ranged between 0.26 and 0.63. The fifth item on this scale had a notably low reliability, 0.02

and 0.06 for boys and girls respectively. The most reliable items were the first three on the "gender stereotyping" scale. Modification indices suggest that some of the items are actually measuring more than one latent factor. In particular, three of the "math utility" items seem to be measuring the students' anxiety as well. Also, responses to item no. 2 on the "math anxiety" scale reflects substantial components of both math utility and gender-stereotyping.

Table 2. Reliability estimates - Analysis 1.

Item	Squared multiple corr Boys	Squared multiple corr Girls	Item	Squared multiple corr Boys	Squared multiple corr Girls
	Math utility			Math anxiety	
mu1	.22	.25	ma1	.34	.35
mu2	.21	.23	ma2	.26	.37
mu3	.39	.35	ma3	.28	.31
mu4	.38	.36	ma4	.58	.63
mu5	.20	.21	ma5	.02	.06
mu6	.31	.34		Gender-stereotyping	
mu7	.32	.46	ms1	.69	.67
mu8	.21	.31	ms2	.74	.67
			ms3	.47	.45
			ms4	.27	.28

Analysis 2.

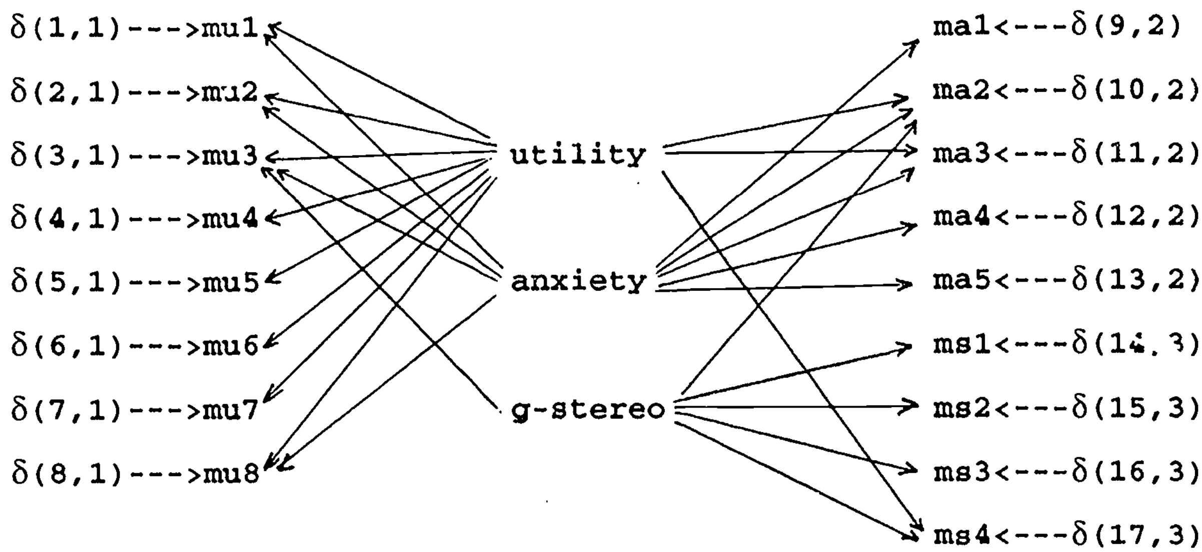
Table 3 illustrates the steps taken in building the model which best fit the data. Allowing mu2, mu3, and mu8 of the "math utility" scale items to load on "math anxiety" and ma2 of the "anxiety" scale to load on "utility", improved the goodness of fit of the model dramatically, i.e. model 2. However, responses to some other items still reflect appreciable components of latent traits other than the ones they are intended to measure. Table 3 shows the steps taken in developing the model which best fit the data. The χ^2-index for the final model, model 4, is 2.8.

Table 3. χ^2-indices for 4 measurement models - Analysis 2.

	χ^2	df	χ^2-index	Improvement
Model 1	879.2	116	7.5	
Model 2	423.5	112	3.8	χ^2_4=455.8
Model 3	332.6	108	3.1	χ^2_4= 90.9
Model 4	297.9	107	2.8	χ^2_1= 34.7

When model 4 is tested on the second half of the boys data, it shows as good a fit as it does on the training data set, with χ^2-index of 3.2. The relationships between observed responses and the latent factors according to model 4 are illustrated in Figure 2.

Figure 2. Fitted model, model 4.



<u>Analysis 3.</u>

<u>Similarity of factor patterns</u>. This hypothesis is tested by specifying an initial model in which both groups have the same factor patterns and starting values, i.e. H_1: same Λ pattern holds for both groups. This model resulted in a χ^2 index of 3.3 which indicates a borderline fit. Standardised residuals are rather large, ranging between -5 and 9.

Examination of estimates of factor loadings for boys and girls reveals some rather large differences in the loadings of some items. For example, estimates of utility loadings ma2, mu7, ms4 are -0.45, -0.92, 0.22 for boys and -0.07, -1.17, 0.06 for girls, respectively. The estimates of anxiety loadings on ma2 is -0.36 for boys and -0.77 for girls, and the g-stereotyping loadings on ma2 is 0.15 for boys and 0.30 for girls. These items should be studied in more detail and checked carefully for differential factor loadings for the two groups.

Equality of factor loadings. To test this hypothesis, a model with equal factor loadings for boys and girls is postulated, i.e. H_2: $\Lambda_{boys}=\Lambda_{girls}$. This model yielded a χ^2 index of 3.2 and standardised residuals ranging between -16 and 42, which leads us to reject the hypothesis of equal factor loadings for boys and girls. A χ^2 index of 2.7 for testing the hypothesis of equality of units of measurement for boys and girls does not provide enough evidence against this hypothesis. The covariances of mu7 and mu1, mu7 and mu5, ma2 and ma1, ma3 and ma2, and ms4 and mu1 have the largest standardized residuals, (greater than 10). These are the same items that were identified in the previous step for possible differential factor loadings.

The increase in χ^2 from that of the model in step 1, can be used to test the hypothesis of equality of units of measurement in the two groups. The increase in χ^2 is 62.52 with 23 degrees of freedom, which gives a χ^2-index of 2.7. Hence the data do not yield enough evidence to reject the hypothesis of equal units of measurement for boys and girls.

Equality of accuracy of measurement. A model is used to test this hypothesis in which the errors of measurement are specified as equal while the factor loadings are not, i.e. H_3: $\Theta_{boys}=\Theta_{girls}$. The factor patterns are defined as similar for the two groups. This model has a χ^2 index of 3.1, and standardised residuals ranging between -9 and 9.

Examination of estimates of factor loadings pointed to the same problem areas as in the above two steps.

VI. Conclusion

Based on the above analyses, it is apparent that most of the items explored are in need of serious revision. The items in general, have poor reliability and they seem to be measuring a mixture of traits. In particular, the item "when I cannot figure out a math problem, I feel lost in a maze" has a near zero reliability. Item no. 3 on the math utility scale is an example of a poorly worded item. "I would like a job that lets me use math", might be soliciting a subjective preference (like/dislike), rather than an objective opinion on the importance of mathematics in today's jobs. Although the items on the math utility scale are intended to tap the student's objective opinions with regard to how useful mathematics is in today's society, the responses to some of the items reflect opinions that are strongly tainted by the student's level of anxiety. For example, a student who had a high level of anxiety toward mathematics, or is unsure of his/her abilities in math, might tend to downplay the importance/usefulness of math in today's society.

The multisample analyses do not provide enough evidence against the hypotheses that boys and girls have the same factor pattern, and the same units of measurement. Factor loadings are found to be different for the two groups.

The following items need further investigation for possible differential loadings of the three factors studied in this paper:

1. Math is not needed in everyday living
2. It scares me to have to take math
3. A woman needs a career just as much as a man does.

Finally, it is the author's recommendation that comparisons among countries on the basis of responses to the attitude questionnaire items should not be considered valid unless it can be demonstrated that these items measure the same traits in the same units with the same accuracy across nations.